Editorial

MICHAEL SEADLE
Dean, Faculty of Arts and Director, Berlin School of Library and Information Science and Editor, Library Hi-Tech, Germany

World Digital Libraries 5(2): v

Incorporating Expectations

A striking feature of five of the six articles in this issue, and indeed a striking feature of most research articles in the world of contemporary Library and Information Science (LIS) research, is the dominant role of statistical information. Often this comprises simple descriptive statistics using raw counts and percentages. Far more rarely do inferential statistics and hypothesis testing come into play, but that will likely change as schools of library and information science include more training in statistics as part of the basic coursework. Statistics in the nineteenth and early twentieth century meant primarily the raw numbers associated with particular information sources. The census is a good example, especially the Indian Census from the 1970s with its volumes and volumes of detailed information about life in the subcontinent. The information was, however, locked up in ink and paper and sociologists and economists who wanted to work with it had to transcribe the data (onto punch cards in those days) in order to run a computer analysis. Today this kind of information is readily available digitally and remains a core element of library content. Nonetheless statistics presents a problem for LIS research that is common to many other academic areas: how to combine the statistical results of multiple studies whose samples, sources, and definitions are similar but not sufficiently comparable to embed the original data in a new meta-analysis. This is a typical problem for election polling organizations, where each polling organization does its own sampling and announces new results based on their latest sample. Because each organization has its own statistical model to determine who is likely to vote and who is likely to be telling the truth when questioned, the results are not readily sharable across organizations, and because the organizations mostly want to offer new information, they tend not to build on their own prior data, even if a good statistical argument for combining data across time could be made. One analyst, Nate Silver, approached the problem differently in his FiveThirtyEight blog in the New York Times: he used Bayesian statistics.

Andrew Hacker (2013) offers a concise explanation of Bayesian statistics:

"The Bayesian approach to probability is essentially simple: start by approximating the odds of something happening, then alter that figure as more findings come in. So it's wholly empirical, rather than building edifices of equations."

In mathematical terms, Bayes Theorem is: "The probability of H conditional on E is defined as $PE(H) = P(H \& E)/P(E)$, provided that both terms of this ratio exist and $P(E) > 0$". (Joyce, 2008) LIS professionals can apply this simple theorem to a wide range of conditions where they have some basis for their expectations and where they wish to predict an outcome. Predicting the acceptance of particular sets of online journals is one example. Rather than doing a new study with a narrow local sample, an LIS professional could take a number of new and previous studies and look at the likelihood of acceptance among students and professors. Expectations are set by older studies. Some newer studies may show widespread acceptance. Others may show a greater reserve toward electronic resources. This variation represents a normal amount of noise based on local conditions and the accident of those participating in a survey. Bayes Theorem helps to smooth out the noise and to give a prediction whose reliability rests not on a single survey and a (relatively) small sample, but on cumulated expectations from multiple samples. Some thought needs to go into calculating these expectations and LIS professionals need to build reliable and transparent models for explaining their choices.

This editorial is not the place to discuss such models or to offer examples. The goal is merely to remind readers that statistical analysis has become a key — almost an essential — tool for scholarship in our field, and that Bayes Theorem offers a much neglected opportunity to reduce the noise and combine results.

References

Hacker, A., 2013. How he got it right. New York Review of Books, 60(1). Available at: New York Review of Books.

Joyce, James, "Bayes' Theorem", The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/bayes-theorem/>